Proceedings of the 10th International Conference on Music Perception and Cognition (ICMPC 10). Sapporo, Japan.
*Ken'ichi Miyazaki, Yuzuru Hiraga, Mayumi Adachi, Yoshitaka Nakajima, and Minoru Tsuzaki (Editors)*

# Computational Model for Automatic Chord Voicing based on Bayesian Network

Tetsuro Kitahara[1,2], Makiko Katsura[2], Haruhiro Katayose[1,2], and Noriko Nagata[1,2]

[1]*CrestMuse Project, CREST, JST, Japan*
[2]*Graduate School of Science and Technology, Kwansei Gakuin University, Japan*
{t.kitahara,katsura,katayose,nagata}@kwansei.ac.jp

## ABSTRACT

We developed a computational model for automatically voicing chords based on a Bayesian network. Automatic chord voicing is difficult because it is necessary to choose extended notes and inversions by taking into account musical simultaneity and sequentiality. We overcome this difficulty by inferring the most likely chord voicing using a Bayesian network model where musical simultaneity and sequeniality are modeled as probabilistic dependencies between nodes. The model represents musical simultaneity as probabilistic dependencies between voicing and melody nodes while it represents musical sequentiality as probabilistic dependencies between current-chord and previous- or following-chord voiding nodes. The model makes it possible to take into account both simultaneity and sequentiality at a single inference process. Experimental results of chord voicing for jazz musical pieces showed that our system generated chord voicings that had appropriate simultaneity and sequentiality.

## I. INTRODUCTION

*Voicing* is the simultaneous vertical placement of notes in order to obtain the effective sound of a chord. Accompaniment parts in popular and jazz music are often represented as a sequence of chord symbols (called a chord progression), not as definite notes. The voicing of chords is therefore an important task in music arrangement.

Only a few attempts at building a computational model for voicing chords have been made. There are a number of studies of generating or analyzing chord progressions [Pachet, 2000, Chemiller, 2001], but they have not dealt with determining the voicings for given chord progressions.

One of the reasons chord voicing is an important research topic is that there are many possible chord voicings for the same chord progression. Given a chord progression, the component notes of each chord are uniquely determined. The following points are, however, not uniquely determined: (1) which inversion is used (i.e., which note is on top, bottom, or in the middle), (2) whether extended notes are added, and (3) whether some notes are omitted. Arrangers deal with these matters taking into account the matching with the melody line (*simultaneity*) and the temporal continuity (*sequentiality*). Because the selection of extended notes directly affects the impression the music has on the audience, especially in jazz music, chord voicing is a nontrivial task.

One possible way to resolve these issues is to use existing music theories. Music theories give us many suggestions for music arrangements including chord voicing. Emura et al. [Emura, 2008] developed a rule-based chord voicing system based on their implementation of music theories. However, because music theories are originally studied for supporting human music activity, they are ambiguous to implement on a compmuter. When they are implemented as a set of rules, for example, some rules may often be conflicted and the priority of such rules are not clear.

In this paper, we propose a computational chord voicing model where the most likely voicings are inferred based on probability theory. Our model represents the two musical aspects, i.e., *simultaneity* and *sequentiality*, as probabilistic dependencies between nodes in a Bayesian network. Simultaneity is represented as dependencies between voicing and melody nodes, while sequentiality is represented as dependencies between current-chord and previous- or following-chord voicing nodes. Because all the dependencies are taken into account when the conditional probabilities of chord-voicing nodes are estimated, our model can infer the most likely voicings that have both simultaneity and sequentiality. In addition, because our model does not use explicit music theories but rather acquires voicing rules implicitly from given performance examples, it can be applied to genres that do not have well-organized theories as long as a number of performance examples are available.

The rest of the paper is organized as follows: Section II details the problems of chord voicing. Section III describes our computational chord voicing model based on a Bayesian network. Section IV reports our experimental results. Section V concludes the paper.

## II. PROBLEMS WITH CHORD VOICING

### A. Statement of Problem

The problem is estimating the most likely voicing for each chord of a given chord progression. We target jazz music to be played on an electronic organ that has one pedal and two manual keyboards. The inputs are the data written on a lead sheet, that is, a melody line and a chord progression. The given melody line is to be played by the right hand. The outputs are the left-hand voicing and bass note (to be played on the pedal keyboard) for each chord of the given chord progression. For simplicity, no passing chords or passing notes in the bass lines

ISBN: 978-4-9904208-0-2 © 2008 ICMPC 10

are added.

### B. Problem

The main difficulty in chord voicing lies in the fact that there are many possible placements of notes for the same chord. The main factors to be considered are summarized as follows:

- *Adding extended notes*
  Adding extended notes, e.g., 9ths, 11ths, and 13ths, is indispensable in order to obtain an effective sound, especially in jazz music. In choosing extended notes, suiting the melody line has to be considered. For example, notes that may produce an unharmonious sound should be avoided.
- *Choosing omitted notes*
  Simply adding extended notes may often create dissonant intervals. For example, the F note on the C11 chord creates the minor ninth interval with the E note. Some notes should be omitted to solve this problem. Notes often need to be omitted due to performance limitations (e.g., the number of fingers of the player).
- *Choosing inversions*
  The appropriate inversion should be determined taking into account the temporal continuity.

Conklin pointed out two aspects of music, *simultaneity* and *sequentiality*, to analyze note-sequence patterns of polyphonic music [Conklin, 2002]. These two aspects are also important in chord voicing. The simultaneity here means that chord voicings should suit the other parts, such as the melody part, and the sequentiality means that each chord voicing should be smoothly connected to the previous and following chord voicings. Roughly speaking, the simultaneity corresponds to extended and omitted notes and the sequentiality corresponds to inversions. In practice, however, these two aspects are mutually dependent; therefor, they should be dealt with in a unified framework.

## III. COMPUTATIONAL MODEL BASED ON BAYESIAN NETWORK

To take into account the above-mentioned aspects when voicing chords, we propose a computational model for chord voicing based on a Bayesian network. A Bayesian network is a probabilistic graphical model that represents a set of nodes and their probabilistic dependencies. A Bayesian network is therefore suitable for modeling the vertical (simultaneous) and horizontal (sequential) dependencies between various elements of music. In fact, it has been used in a musical scene analysis system called OPTIMA [Kashino, 1998]. The Bayesian network used by OPTIMA, however, cannot be applied to chord voicing without modification because it aims to form a symbolic representation of a given musical audio signal. We therefore build a Bayesian network that can be used for chord voicing.

### A. Basic Policies

As described above, we should take into account both simultaneity and sequentiality in designing a chode voicing model. We discuss how to consider simultaneity and sequentiality as follows:

- *Simultaneity*
  In our case, simultaneity represents the relationship between the melody and voicing. Similarly to studies of harmonization based on hidden Markov models [Kawakami, 2000], we consider an observed melody line to be generated from the harmony including the voicings.
- *Sequentiality*
  Sequentiality here means the temporal continuity of the voicings of successive chords. In particular, the contiuity of the top and bottom notes are important. The chord voicing model should therefore separately represent the temporal continuity of the top, bottom, and other notes in chord voicings. In addition, when arrangers choose the voicing of a certain chord, they take into account whether this voicing can be smoothly connected to the following chord. In other words, the voicing of a certain chord depends not only on that of the previous chord but also that of the following chord. The chord voicing model should therefore evaluate the influence of the voicing of the target chord on that of the following chord.

In the rest of this section, we describe a left-hand voicing model and a bass model that we designed on the basis of these basic ideas. Ideally, left-hand chord voicing and bass note determination should be performed in a unified framework because they are not independent. However, to reduce the computational cost, we build separate models for left-hand chord voicing and bass note determination. Integrating the two models will be future work.

### B. Left-hand Voicing Model

**Figure 1** shows our left-hand voicing model. This model has three layers: chord-name, chord-voicing, and melody.

The chord-voicing layers have nodes for the current, previous, and following chords. The chord-voicing nodes for each chord are divided into three nodes: top, middle, and bottom. If a voicing of the chord CM7 is C-E-G-B, the top, middle, and bottom nodes are set to “B”, “EG”, and “C”, respectively. The reason the top and bottom notes are treated separately from the others is that the sequentiality of these notes is particularly important. The top nodes of successive chords are connected by arcs. That means that the top nodes of the following chord depends on that of the current chord, which depends on that of the previous chord. The middle and bottom nodes of successive chords are connected in the same way. These connections represent the horizontal (sequential) dependencies of chord voicing.

The chord-name layers represent the chord names taken from an input lead sheet. The chord-name nodes for the current and following chords are prepared, and they are connected to the chord-voicing nodes for the current and following chords, respectively. Because the voicing for the previous chord has been determined, the chord-name node for the previous chord is not prepared.

The melody layers have 12 nodes corresponding to all the

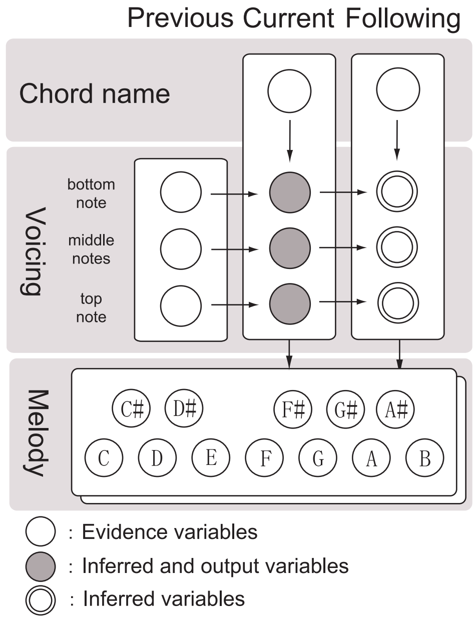


**Figure 1: Left-hand voicing model.**

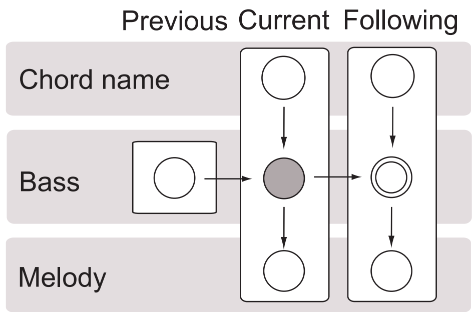


**Figure 2: Bass model.**

pitch classes. Each node represents the relative length of the appearance of the corresponding pitch class, which is discretized to $\{0.0, 0.2, 0.4, \cdots, 1.0\}$ to reduce the computational cost. If the melody node for a certain pitch class has a high value, it should be seriously avoided for some note in a voicing to create a dissonant interval combined with this pitch class. If that has a value close to zero, this pitch class could be used mainly as passing notes, so the priority of avoiding a dissonant interval becomes relatively low. Each melody node is connected from every chord-voicing node for the current chord. This connection represent the vertical (simultaneous) dependencies.

### C. Bass Model

**Figure 2** shows our bass model. Similarly to the left-hand voicing model, the bass model has three layers: chord-name, bass, and melody. The bass nodes are prepared for the current, previous, and following chords. The bass nodes for successive chords are connected, which represents the horizontal dependencies of the bass line. The bass nodes and the melody nodes are connected, which represents the vertical dependencies.

**Table 1: Evaluation of musical simultaneity by music experts.**

| | # of chords | Percentage |
|---|---|---|
| Beautiful | 82 | 9.0% |
| Good | 719 | 78.8% |
| Acceptable | 63 | 6.9% |
| Unacceptable | 48 | 5.3% |

### D. Application to Automatic Chord Voicing

As described above, our models are applied to each chord from the beginning to the end of a given chord progression by shifting the current chord. Specifically, the following steps are performed:

1. Set the current chord to the first chord of a given lead sheet.
2. Set the chord-name and melody nodes to values obtained from the lead sheet.
3. Calculate the conditional probabilities for the other nodes.
4. Determine the voicing for the current chord based on the calculated probabilities.
5. Change the current chord to the following chord.
6. Set the chord-voicing nodes for the previous nodes to the determined voicing.
7. Return to 2. until reaching the end.

For simplicity, the pitch ranges for left-hand voicing and bass notes are restricted to C3–A#4 and C2–G3, respectively. The actual pitches of voicing and bass lines are determined within these ranges.

## IV. EXPERIMENTS

### A. Evaluation by Music Experts

We asked experts in jazz music to evaluate the results of our automatic chord voicing model from the viewpoints of musical simultaneity and sequentiality. For simultaneity, three subjects categorized the quality of the voicing of each chord as either beautiful, good, acceptable, or unacceptable. Five jazz musical pieces were used: *As Time Goes By*, *I Left My Heart In SanFrancisco*, *My One And Only Love*, *Misty*, and *Satin Doll*. The number of chords was 304 in total. The training data were taken from 30 jazz musical pieces arranged for electronic organ. One example of the results of chord voicing is shown in **Figure 3**. For sequentiality, two subjects categorized the continuity of each chord change as either good, acceptable, or unacceptable. The number of chord changes was 299 in total.

The results of evaluating musical simultaneity is listed in **Table 1**. The subjects determined 94.7% of chords to be at least acceptable. This means we successfully developed a practical chord voicing model. One possible reason about 5% of chords were unacceptable is that the left-hand voicing and bass models were separate. Developing a unified model that incorporates both left-hand voicing and bass note determination is needed to improve voicing further.

The results of categorizing musical sequentiality are listed in **Table 2**. 84.1% of chord changes were determined to be at least acceptable. One possible reason why 15.9% of chord


**Figure 3: Example of results of chord voicing (Misty)**

**Table 2: Evaluation of musical sequentiality by music experts.**

| | # of chord changes | Percentage |
|---|---|---|
| Good | 180 | 25.8% |
| Acceptable | 407 | 58.3% |
| Unacceptable | 111 | 15.9% |

changes were unacceptable was that the uniform restriction of pitch ranges caused unnecessary pitch motions.

## V. CONCLUSION

In this paper, we proposed a computational model for automatically voicing chords based on a Bayesian network. Using the Bayesian network, we achieved effective chord voicing taking into account both two aspects of music, simultaneity and sequentiality.

Because Bayesian networks are flexible models, various tasks concerning music analysis, composition, and arrangement can also be achieved based on Bayesian networks. Future work will therefore include construction of a unified framework for performing such tasks based on Bayesian networks.

## REFERENCES

[Pachet, 2000] F. Pachet: Computer Analysis of Jazz Chord Sequences. Is Solar a Blues?, *Readings in Music and Artificial Intelligence*, Harwood Academic Publishers (2000).

[Chemiller, 2001] M. Chemiller: Improving Jazz Chord Sequences by Means of Formal Grammars, *Proceedings of the International Computer Music Conference* (2001).

[Emura, 2008] N. Emura, M. Miura, and M. Yanagida: A modular system generating Jazz-style arrangement for a given set of a melody and its chord name sequence, *Acoustical Science and Technology*, Vol. 29, No. 1, pp. 51–57 (2008).

[Conklin, 2002] D. Conklin: Representation and discovery of vertical patterns in music, *Music and Artificial Intelligence: Lecture Notes in Artificial Intelligence*, Vol. 2445, pp. 32–42 (2002).

[Kashino, 1998] K. Kashino, K. Nakadai, T. Kinoshita and H. Tanaka, Application of the Bayesian Probability Network to Music Scene Analysis, *Computational Auditory Scene Analysis*, D. F. Rosenthal and H. G. Okuno (eds.), Lawrence Erlbaum Associates, pp.115–137 (1998).

[Kawakami, 2000] T. Kawakami, M. Nakai, H. Shimodaira, and S. Sagayama, Hidden Markov Model Applied to Automatic Harmonization of Given Melodies, *IPSJ SIG Notes*, 99-MUS-34, pp.59–66 (2000) (in Japanese).